# Why the Symbol Grounding Problem Matters in Virtual Reality: A Meme-Focused Solution Based on the Model Human Processor with Real-Time Constraints

Muneo Kitajima
*Nagaoka University of Technology*
Nagaoka, Niigata, Japan
Email: mkitajima@kjs.nagaokaut.ac.jp

Makoto Toyota
*T-Method*
Chiba, Japan
Email: pubmtoyota@mac.com

Katsuko T. Nakahira
*Nagaoka University of Technology*
Nagaoka, Niigata, Japan
Email: katsuko@vos.nagaokaut.ac.jp

***Abstract*— Virtual Reality (VR) environment presents objects that the users perceive and interact with. The system then transitions to a new state, reflecting the outcome of the interaction based on the user's perception of the presented objects. For seamless interaction between a Virtual Reality (VR) environment and its users, the meanings assigned to objects by VR application developers must align with the meanings perceived by users. This study proposes a method to enable such seamless interaction by treating the objects presented in VR as symbols and addressing the relationship between the meanings embedded in those symbols and the meanings interpreted by users. This is achieved through the lens of the symbol grounding problem, a well-known challenge in the field of artificial intelligence. Based on the Mode Human Processor with Realtime Constraints (MHP/RT), a cognitive architecture capable of handling action selection in everyday environments, we focus on the idea that users' action selections are influenced by memes—culturally transmitted units of meaning passed down through generations. These memes shape users' understanding of their surrounding world. We suggest that the symbol grounding problem can be addressed by identifying and observing the memes that users draw upon in their daily behavior. We suggest that implementing extracted memes within a Virtual Reality (VR) system can address the symbol grounding problem and enable seamless interaction based on a shared understanding of the symbols mediating between the user and the system.**

***Keywords- the symbol grounding problem; meme; virtual reality; artificial intelligence; Model Human Processor with Real-Time Constraints; structured meme theory.***

## I. INTRODUCTION

This article is based on the previous work originally presented in AIVR 2025 [1]. In Section IV, a new Section IV-A is introduced to discuss the process through which "symbols" are generated and structured as memes—a necessary consideration when addressing the symbol grounding problem. Furthermore, Section IV-C explains how securely grounded symbols are utilized in real life contexts and discusses the observability of structured memes.

Humans acquire information about the external world through the five senses and select and execute appropriate actions based on the situation by engaging the Perceptual, Cognitive, and Motor (PCM) processes. In the perceptual process, perceived information is encoded and represented as symbols, enabling thought to occur in the cognitive process. In the thinking process, memory is used to successively transform the symbols into new representations. Part of the outcome of cognition is a representation of an action sequence that can be executed through the motor process. Building on the idea that human intelligence arises from thought processes that manipulate symbols, Newell proposed the Physical Symbol System as theoretical framework for understanding human intelligence [2]. This idea laid the foundation for Soar [3], [4], one of the most well-established cognitive architectures.

The information received by the perceptual process through the sensory organs originates in the real world. Guided by the sequence of actions represented by symbols generated through the PCM processes, actions are executed in the real world, leading to changes in the environment. When attempting to artificially replicate the interactions humans perform in the real world using a physical symbol system, the system must be capable of linking symbolic representations to real-world referents and deriving meaningful understanding from its interactions with the environment.

Achieving this capability is a fundamental challenge in Artificial Intelligence (AI) research, known as the Symbol Grounding Problem (SGP) [5]. The SGP concerns a machine's ability to link its symbolic representations to real-word referents and derive meaningful understanding through its interaction with its environment. Specifically, it addresses how machines can understand and represent the meaning of objects, concepts, and events in the real world. Without the ability to ground symbolic representations in the real world, machines cannot acquire the rich and complex meanings needed for intelligent behavior such as language processing, image recognition, and decision-making. Addressing the SGP is crucial to building machines that can perceive, reason, and act like humans.

Virtual Reality (VR) environments represent one of the contexts in which humans engage in interaction. In such environment, users interact with computer-generated three-dimensional visual scenes or multi-sensory simulations created through modeling and simulation. VR applications immerse the users in a computer-generated environment that closely replicate aspects of the real world. In a VR environment, user interaction proceeds through user-perceivable objects provided by VR applications. The meaning that the user gives to the objects generated by the VR applications determines how the user interacts with the objects. VR applications can achieve seamless interaction by appropriately handling the meanings users assign to virtual objects. This shows that the SGP is central to seamless VR environments. This paper proposes a method to address the SGP in VR.

This paper is organized as follows. Section II introduces a

2025, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

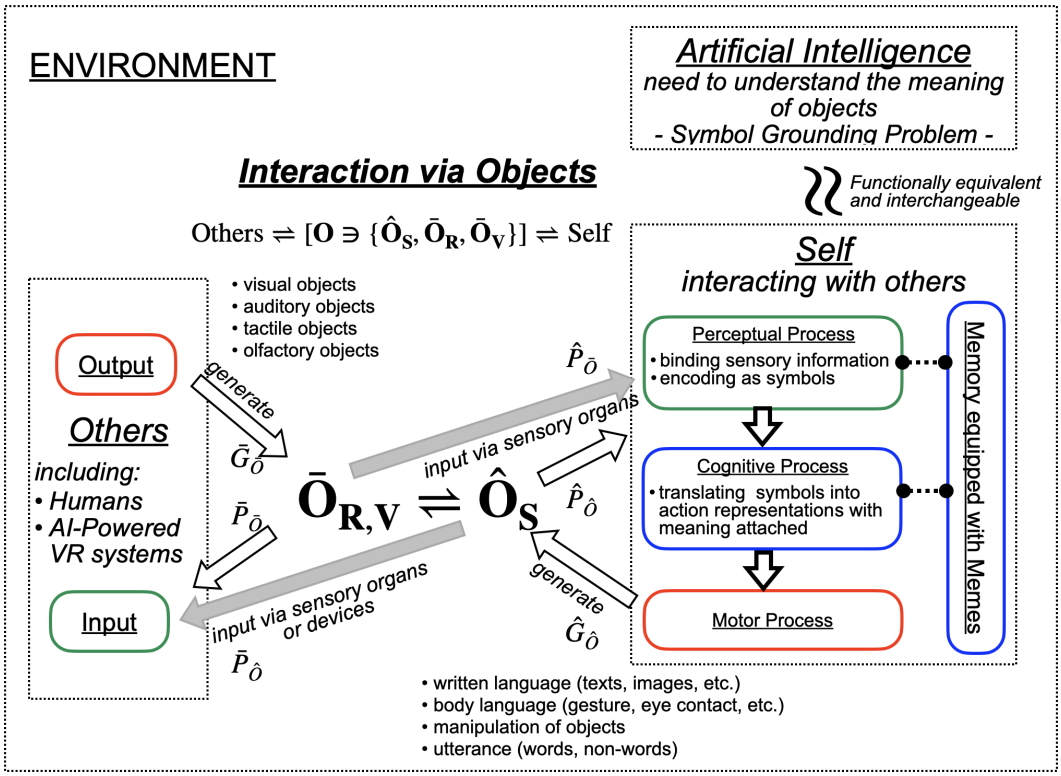


Figure 1. Interaction via objects (adapted from [1, Figure 1])

framework for object-based interaction. Section III examines self–object interaction, approaches to capturing the SGP in AI from an interaction perspective, SGP in AI from the interaction perspective, and how to capture the interaction between self and VR. Section IV proposes methods for generating VR environments that support symbol grounding. Section V concludes this paper by summarizing its findings and outlining future research directions.

## II. Interaction via Objects

This section outlines aspects of individual human interaction through objects, with classifications provided in Section II-A and the PCM processes and memory functions described in Section II-B.

### A. Interaction Types

Individuals live their daily lives interacting with several objects that exist around them. Objects are classified based on whether the self interacts with them directly or indirectly, and by who created them. There are three types of objects defined as follows:

- $\hat{\mathbf{O}}_\mathbf{S}$: Real objects that are directly generated by self, e.g., utterance, written text, drawings, gestures, modeling, etc.
- $\bar{\mathbf{O}}_\mathbf{R}$: Real objects generated by other human beings with whom self is directly interacting.
- $\bar{\mathbf{O}}_\mathbf{V}$: Objects generated by a machine or other human beings with whom self is not directly interacting.

These three types of objects are collectively represented as $\mathbf{O}$ ( $\ni \{\hat{\mathbf{O}}_\mathbf{S}, \bar{\mathbf{O}}_\mathbf{R}, \bar{\mathbf{O}}_\mathbf{V}\}$). In Figure 1, $\mathbf{O}$ is placed in the center, "Self" interacting with $\mathbf{O}$ on its right side, and "Others" interacting with $\mathbf{O}$ on its left side. On top of "Self" is an "Artificial Intelligence" that is functionally equivalent to self and can be replaced. Self operates the PCM processes to generate the objects $\hat{\mathbf{O}}_\mathbf{S}$ ($\hat{G}_{\hat{O}}$). Self also perceives them ($\hat{P}_{\hat{O}}$). The objects self perceives include the objects $\bar{\mathbf{O}}_{\mathbf{R},\mathbf{V}}$ generated by others ($\hat{P}_{\bar{O}}$). Meanwhile, others generate the objects $\bar{\mathbf{O}}_{\mathbf{R},\mathbf{V}}$ ($\bar{G}_{\bar{O}}$). If the other is a human being, they run the PCM process, which is equivalent to the one self runs, to generate the objects $\bar{\mathbf{O}}_\mathbf{R}$ ($\bar{G}_{\bar{O}_R}$). Machines run their own generative mechanisms to produce the objects $\bar{\mathbf{O}}_\mathbf{V}$ as output ($\bar{G}_{\bar{O}_V}$). The input to the others are the objects $\bar{\mathbf{O}}_{\mathbf{R},\mathbf{V}}$ generated by the others themselves ($\bar{P}_{\bar{O}}$), or the objects $\hat{\mathbf{O}}_\mathbf{S}$ generated by self ($\bar{P}_{\hat{O}}$). In summary, interaction via objects can be expressed as follows:

$$\text{Others} \rightleftharpoons [\mathbf{O} \ni \{\hat{\mathbf{O}}_\mathbf{S}, \bar{\mathbf{O}}_\mathbf{R}, \bar{\mathbf{O}}_\mathbf{V}\}] \rightleftharpoons \text{Self}$$

### B. Interaction between Self and $\mathbf{O}$

Interaction between the self and object $\mathbf{O}$ is driven by the PCM processes, which operate using memory processes that are also updated through PCM execution. This section provides an overview of the PCM and memory processes based on the Model Human Processor with Real-Time Constraints (MHP/RT), a cognitive architecture that can simulate everyday action selections [6]–[8].

2025, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

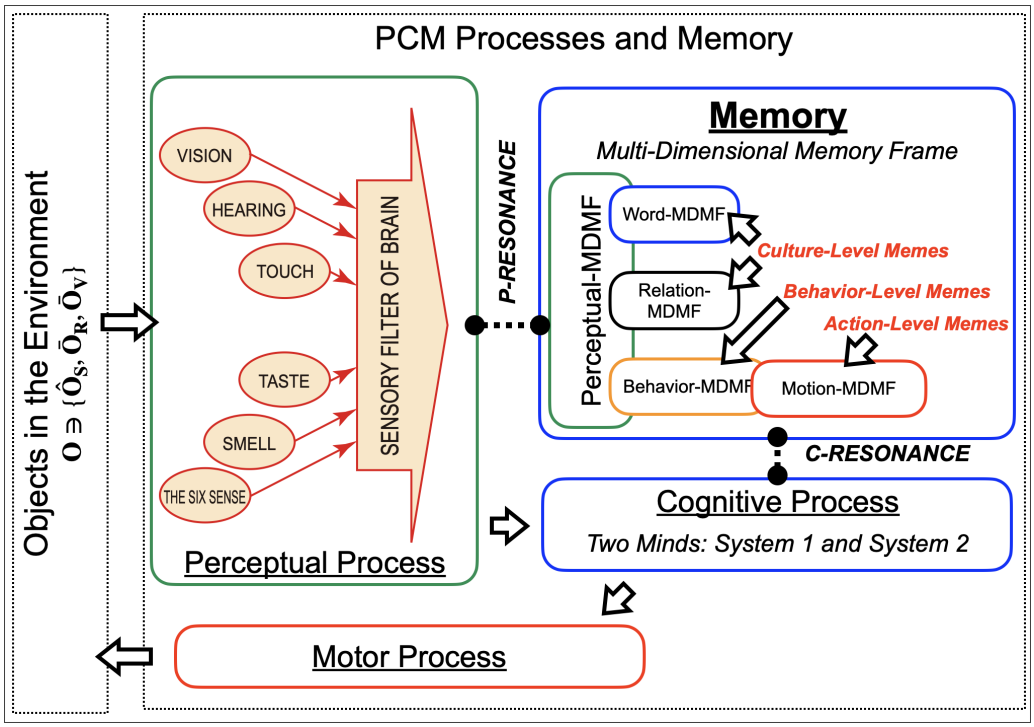


Figure 2. PCM process and memory (adapted from [1, Figure 2])

*1) PCM Process:* When interacting with objects in the environment, humans receive physical and chemical stimuli through sensory nerves at the body's interface with the environment, allowing environmental information to be processed internally. The brain acquires environmental information concerning the current activity of the self through the multiple sensory organs. Further, it generates bodily movements that are suitable for the current environment. The stable and sustainable relationship between the environment and the self is established through continuous coordination between the activity of the self and the resultant changes in the environment, which should affect the self's next action.

Figure 2, adapted from [9, Figure 1] with some modifications, shows the PCM process, based on the MHP/RT cognitive architecture [7], [8], by which environmental information is taken into the body via sensory nerves, processed in the brain, and then acted upon by the external world via motor nerves. This process uses memory, which is modeled as the Multi-Dimensional Memory Frame. It comprises the Perceptual-, Behavior-, Motor-, Relation-, and Word-Multi-Dimensional Memory Frame. The Perceptual-Multi-Dimensional Memory Frame overlaps with the Behavior-, Relation-, and Word-Multi-Dimensional Memory Frame. This is the unique and indispensable configuration of memories defined by the Multi-Dimensional Memory Frame for spreading activation from the Perceptual- to Motor-Multi-Dimensional Memory Frame, which connects perception with bodily movements.

Perceptual information taken in from the environment through sensory organs *resonates* with information in the Multi-Dimensional Memory Frame, which is called P-Resonance [9]. In Figure 2, this process is indicated by the symbol •—•. Resonance occurs first in the Perceptual-Multi-Dimensional Memory Frame and activates the memory network. After that, the activation spreads to the memory networks that overlap with the Perceptual-Multi-Dimensional Memory Frame, and finally to the Motor-Multi-Dimensional Memory Frame. In cognitive processing based on the Two Minds framework [10], [11], conscious processing (System 2) and unconscious processing (System 1) operate in an interrelated manner [9], [12]. System 2 utilizes the Word- and Relation-Multi-Dimensional Memory Frame via C-Resonance, while System 1 draws on the Behavior- and Motor-Multi-Dimensional Memory Frame via the same mechanism. Motor sequences are then expressed according to the Motor-Multi-Dimensional Memory Frame. The memories involved in the production of actions are updated to reflect the traces of their use process and influence the future action selection process.

*2) Memory and Memes:* When the PCM process is running, the contents of Perceptual-Multi-Dimensional Memory Frame are updated in response to the perceptual process, those of Word-, Relation-, and Behavior-Multi-Dimensional Memory Frame are updated in response to the cognitive process, and those of Motor-Multi-Dimensional Memory Frame are updated in response to the motor process. Figure 2 illustrates the memories involved in the PCM process—collectively referred to as the Multi-Dimensional Memory Frame—as traces of its operation, and classifies them into five subtypes: Perceptual-, Word-, Relation-, Behavior-, and Motor-Multi-Dimensional

2025, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

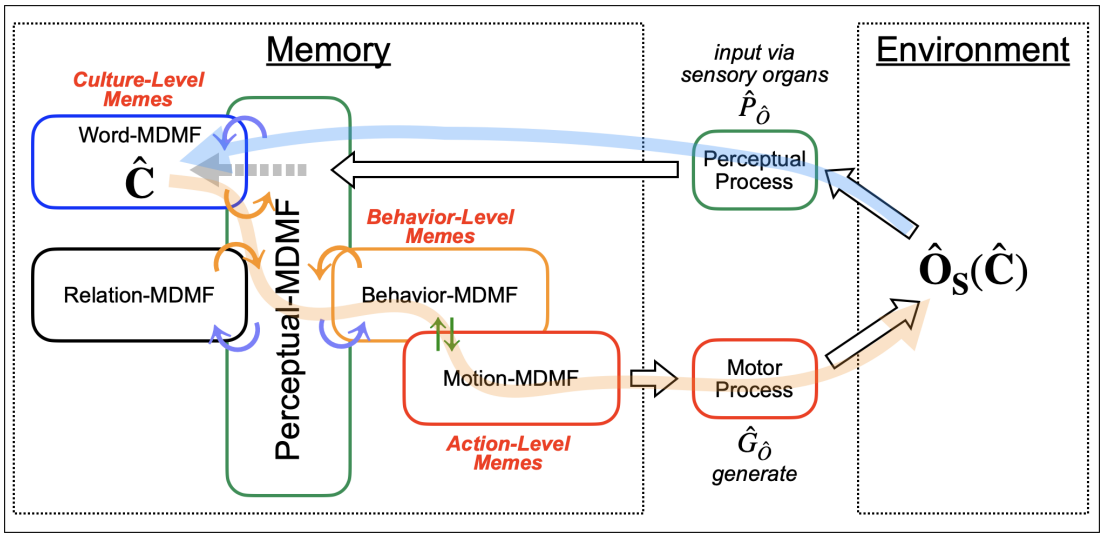


Figure 3. Symbol grounding when the self generates an object $\hat{\mathbf{O}}_{\mathbf{S}}(\hat{\mathbf{C}})$ that embodies the concept $\hat{\mathbf{C}}$ (adapted from [1, Figure 3])

Memory Frame. Specifically, this is an expression of the way in which the memories are structured, focusing on the continuous updating of memory associated with the execution of PCM process.

Alternatively, the memory system can be viewed from the perspective of memory use. The integrated sensory information first activates the Perceptual-Multi-Dimensional Memory Frame; then the activation spreads to the Word-, Relation-, and Behavior-Multi-Dimensional Memory Frame, and finally to the Motor-Multi-Dimensional Memory Frame bound to the motor nerves. Imitation—doing what one observes—forms the basis of behavior. Therefore, behaviors that can be imitated across generations are preserved as sustainable behaviors. Therefore, we can organize the Multi-Dimensional Memory Frame, which is used by the PCM processes and updated by their execution, regarding *memes* that can be inherited across generations [13].

Word is considered the archetype of meme [14]. Words—also referred to as symbols—are gradually integrated into the environment in stages. First, they appear as thesauruses, which are structured lists of words grouped by synonyms and related concepts. Next, they are incorporated as individual languages used for person-to-person communication, which may involve both literal and metaphorical usage. Finally, they are embedded in cultural languages, where successful communication depends on a shared understanding of the common sense established within a specific community. These three forms circulate among individuals and persist from generation to generation [15].

Thesauruses, individual languages, and cultural languages increase in complexity in that order, based on the patterns through which they are linked to objects in the environment. Thesauruses are associated with the objects in the environment that are encoded in the neural networks in the initial development stage from birth to the age of 3 years. Individual languages are associated with the objects in the environment and the symbols that have already been incorporated in the environment. The same is true for cultural languages. In the Structured Meme Theory (SMT) [13], which is the memory construct to be used and updated by the MHP/RT, the patterns that represent the thesauruses, individual languages, and cultural languages are called action-level meme (A-meme), behavior-level meme (B-meme), culture-level meme (C-meme), respectively.

As shown in the rounded corner rectangle "Memory" in Figure 2, the relationships between the three levels of memes and the Multi-Dimensional Memory Frame are described as follows:

- A-memes represent bodily actions stored in the Motor-Multi-Dimensional Memory Frame.
- B-memes represent behaviors in the environment stored in the Behavior-Multi-Dimensional Memory Frame.
- C-memes represent culture stored in the Relation- and Word-Multi-Dimensional Memory Frame.

## III. Details of Interaction between Self and O

This section describes how individual human PCM processes and memory work for each of the interaction types introduced in Section II, and clarifies how the symbol grounding problem is involved and solved.

### A. *Interaction between Self and* $\hat{\mathbf{O}}_{\mathbf{S}}$

The process of generating an object $\hat{\mathbf{O}}_{\mathbf{S}}(\hat{\mathbf{C}})$ that represents the concept $\hat{\mathbf{C}}$ by self ($\hat{G}_{\hat{O}(\hat{C})}$) and perceiving the generated result ($\hat{P}_{\hat{O}(\hat{C})}$) is shown in Figure 3 using Figure 2 as a basis. According to this process, an object that matches the concept is generated. Figure 4 shows the flow of activation within the Multi-Dimensional Memory Frame involved in this process.

This object generation process is indicated along the thick orange arrow in Figure 3. It is denoted symbolically as follows:

2025, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

1) The concept $\hat{\mathbf{C}}$ is represented as a symbol within C-memes, and the nodes associated with it in the Word-Multi-Dimensional Memory Frame are activated.
2) Activation spreads from the Word- to Perceptual-Multi-Dimensional Memory Frame.
3) Activation spreads from the Perceptual- to Relation-Multi-Dimensional Memory Frame that encodes C-memes, and reaches Motor-Multi-Dimensional Memory Frame encoding A-memes via Behavior-Multi-Dimensional Memory Frame encoding B-memes.
4) According to the contents of activated Motor-Multi-Dimensional Memory Frame, the motor process operates to express bodily movements, which result in the generation of object $\hat{\mathbf{O}}_{\mathbf{S}}(\hat{\mathbf{C}})$ in the environment.

Figure 4. Object generation process (adapted from [1, Figure 4])

Generation Path [G-SS]

$$\hat{\mathbf{C}} \Rightarrow \left[ \frac{\text{W-, R-, B-MDMF}}{\text{P-MDMF}} \rightleftharpoons \text{M-MDMF} \right] \Rightarrow \hat{\mathbf{O}}_{\mathbf{S}}(\hat{\mathbf{C}}) \quad (1)$$

Here, the part indicated by $[\cdots]$ in the middle of the expression shows the process of spreading activation in the Multi-Dimensional Memory Frame from the viewpoint of the activated location in the memory. The symbol [G-SS] reads as follows; [**G**enerate - symbol in **S**elf via memory of **S**elf].

The generation process [G-SS] can be rearranged from the perspective of memes, i.e., memories of contents, as follows:

meme mapping

$$\hat{\mathbf{C}} \Rightarrow \left[ \frac{\text{C-memes} \rightleftharpoons \text{B-memes}}{\text{P-MDMF}} \rightleftharpoons \text{A-memes} \right]_{\text{M-S}} \Rightarrow \hat{\mathbf{O}}_{\mathbf{S}}(\hat{\mathbf{C}}) \quad (2)$$

The generated object $\hat{\mathbf{O}}_{\mathbf{S}}(\hat{\mathbf{C}})$ shown on the right end is associated with a state in which the C-, B-, and A-memes activated in the process of spreading activation in the Multi-Dimensional Memory Frame starting from the concept $\hat{\mathbf{C}}$ shown on the left end. The association is represented in the middle, $[\cdots]_{\text{M-S}}$, called "meme mapping" from $\hat{\mathbf{C}}$ to $\hat{\mathbf{O}}_{\mathbf{S}}(\hat{\mathbf{C}})$.

1) $\hat{\mathbf{O}}_{\mathbf{S}}(\hat{\mathbf{C}})$ is perceived and the activation spreads within the Perceptual-Multi-Dimensional Memory Frame.
2) The activation spread from the Perceptual- to Word-Multi-Dimensional Memory Frame results in the activation of symbol related to the perceptual representation.

Figure 5. Object recognition process (adapted from [1, Figure 5])

Figure 5 shows the object recognition process. The thick blue arrow in Figure 3 indicates this process. It is denoted symbolically as follows:

Recognition Path [R-SS]

$$\hat{\mathbf{O}}_{\mathbf{S}}(\hat{\mathbf{C}}) \Rightarrow \left[ \frac{\text{P-MDFM}}{\text{W-MDMF}} \right] \Rightarrow \hat{\mathbf{C}} \quad (3)$$

The symbol [R-SS] reads [**R**ecognize object - generated by **S**elf using memory of **S**elf]. In [R-SS], if the concept $\hat{\mathbf{C}}$ is strongly activated at the end of the recognition path, then $\hat{\mathbf{O}}_{\mathbf{S}}(\hat{\mathbf{C}})$ correctly realizes $\hat{\mathbf{C}}$ in the real world. In this case, [G-SS] and [R-SS] are connected and closed, and the symbol $\hat{\mathbf{C}}$ and $\hat{\mathbf{O}}_{\mathbf{S}}(\hat{\mathbf{C}})$ are cognitively replaceable, which is represented by $\hat{\mathbf{C}} \equiv \hat{\mathbf{O}}_{\mathbf{S}}(\hat{\mathbf{C}})$. This state can be regarded as a state in which symbol grounding has been achieved within self as shown by Figure 6.

- Symbol $\hat{\mathbf{C}}$ in C-memes activates B- and A-memes via the Perceptual-Multi-Dimensional Memory Frame to generate $\hat{\mathbf{O}}_{\mathbf{S}}(\hat{\mathbf{C}})$.
- Perception of $\hat{\mathbf{O}}_{\mathbf{S}}(\hat{\mathbf{C}})$ activates $\hat{\mathbf{C}}$ and its associated activation pattern in the Multi-Dimensional Memory Frame.
- The perceptual representation of $\hat{\mathbf{O}}_{\mathbf{S}}(\hat{\mathbf{C}})$ in the Perceptual-Multi-Dimensional Memory Frame is associated with $\hat{\mathbf{C}}$ in Word-Multi-Dimensional Memory Frame.
- In the future, $\hat{\mathbf{C}}$ in the Word-Multi-Dimensional Memory Frame activates perceptual representation of $\hat{\mathbf{O}}_{\mathbf{S}}(\hat{\mathbf{C}})$ even if it does not exist in the real world, which enables the self to perceptually simulate the concept along with the activation of the Multi-Dimensional Memory Frame necessary to actually generate the object.

Figure 6. Symbol grounding of concept $\hat{\mathbf{C}}$ in self (adapted from [1, Figure 6])

### *B. Symbol Grounding Problem in Artificial Intelligence*

The symbol grounding problem in AI shown in Figure 1 is solved by the fact that an activation pattern equivalent to the activation pattern of memes in the Multi-Dimensional Memory Frame that occurs in the process of generating self's $\hat{\mathbf{O}}_{\mathbf{S}}$ also occurs in AI. By ensuring that the meme mappings occur within AI and self are equivalent, i.e., $[\cdots]_{\text{M-AI}} \equiv [\cdots]_{\text{M-S}}$, AI can be a substitute for self. This is summarized in Figure 7, where "Perceptual-Information-Encoding-in-AI" is the substitute for the Perceptual-Multi-Dimensional Memory Frame of human to perform A/D transformation to input the real world data to the AI system. Since the memes are knowledge passed down from generation to generation, it is considered possible to represent them by symbols. The Perceptual-Information-Encoding-in-AI can also be represented in symbols by encoding environmental information by sensors that perform the same function as sensory organs. The symbol grounding problem in AI is thought to be solved by elucidating memes and implementing them in AI. Methods for elucidating memes are discussed in Section IV-C.

2025, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

1) The symbol $\hat{\mathbf{C}}$, which is common to self's, in C-memes activates the Perceptual-Information-Encoding-in-AI as well as the associated C-memes in AI.
2) B-memes in AI are activated via the activated portion of Perceptual-Information-Encoding-in-AI.
3) The part of A-memes in AI that overlap with the activated B-memes in AI is activated.

*The steps 1, 2, and 3 constitute the meme-mapping of AI, i.e.,* $[\cdots]_{M\text{-}AI}$.

4) What is expressed by the activated A-memes in AI is implemented as $\hat{\mathbf{O}}_{\mathbf{AI}}(\hat{\mathbf{C}})$ in the real world via appropriate actuators.
5) Upon input of the object $\hat{\mathbf{O}}_{\mathbf{AI}}(\hat{\mathbf{C}})$ in AI, activation spreads in the Perceptual-Information-Encoding-in-AI, followed by the activation of the symbol $\hat{\mathbf{C}}$ in C-memes in AI.
6) The part in the Perceptual-Information-Encoding-in-AI that corresponds to $\hat{\mathbf{O}}_{\mathbf{S}}(\hat{\mathbf{C}})$ and the symbol $\hat{\mathbf{C}}$ integrate the C-, B-, and A-memes in AI activated in the steps 1, 2, and 3 to form an integrated association. At this point, $\hat{\mathbf{C}} \equiv \hat{\mathbf{O}}_{\mathbf{AI}}(\hat{\mathbf{C}})(= \hat{\mathbf{O}}_{\mathbf{S}}(\hat{\mathbf{C}}))$ is established, in other words, the symbol $\hat{\mathbf{C}}$ both the AI and the self commonly recognize has been grounded, guaranteed by the relationship $[\cdots]_{\text{M-AI}} \equiv [\cdots]_{\text{M-S}}$.

Figure 7. Symbol grounding of concept $\hat{\mathbf{C}}$ in AI (modified from [1, Figure 7])

### C. Interaction between Self and $\bar{\mathbf{O}}_{\mathbf{R}}$

Consider the case where the other human generates an object. The object generation process for the concept $\bar{\mathbf{C}}$ performed by the other human is represented as $\bar{\mathbf{C}} \Rightarrow [\cdots] \Rightarrow \bar{\mathbf{O}}_{\mathbf{R}}(\bar{\mathbf{C}})$, which is a the-other-human's version of [G-SS]. The other human spreads activation in his/her own Multi-Dimensional Memory Frame. The meme mapping used for the other human is denoted as $[\cdots]_{\text{M-O}}$. The self interacting with the other human who has just generated $\bar{\mathbf{O}}_{\mathbf{R}}(\bar{\mathbf{C}})$, which could be perceivable objects such as a gesture, utterance, or drawing, perceives and recognizes it according to the following path:

Recognition Path for the Object Generated by Other [R-OS]

$$\bar{\mathbf{O}}_{\mathbf{R}}(\bar{\mathbf{C}}) \Rightarrow \left[\begin{array}{c}\text{P-MDFM}\\ \hline \text{W-MDMF}\end{array}\right] \Rightarrow \hat{\mathbf{C}} \tag{4}$$

If the symbol $\bar{\mathbf{C}}$ $(\equiv \bar{\mathbf{O}}_{\mathbf{R}}(\bar{\mathbf{C}}))$ held by the other human and the symbol $\hat{\mathbf{C}}$ held by the self are identical, the symbol is successfully transmitted through the object expressed by the other. For example, the other person holds a certain word $\bar{\mathbf{C}}$ in his/her mind and expresses it physically through gestures, and then the self sees it and assigns the word $\hat{\mathbf{C}}$ to it. The latent word of the other is connected to the self's latent word through the physical actions of the other person. Consider the case of communication via words, where the self and the other look at a sequence of words $\tilde{\mathbf{C}}$. The self and the other perform symbol grounding according to their respective generation paths; $[\cdots]_{\text{M-S}}$ and $[\cdots]_{\text{M-O}}$ are included in each symbol grounding process. If the self and the other have grown up in the same environment, which is the necessary condition for them to have a common set of memes, then the relation $[\cdots]_{\text{M-S}} \equiv [\cdots]_{\text{M-O}}$ would hold, and the shared symbols have the same meaning. However, in the case of $[\cdots]_{\text{M-S}} \neq [\cdots]_{\text{M-O}}$, the meaning of all visually shared symbols may not be shared. For example, the phrase "see you on the ground floor" may trigger different behaviors depending on the culture to which the reader of the phrase belongs.

### D. Interaction between Self and $\bar{\mathbf{O}}_{\mathbf{V}}$

In interactions between the self and other humans, the interaction is symmetric because both parties are human. Specifically, in the section represented by $[\cdots]$, which connects symbols and objects along the generation and recognition pathways, activation spreads within the Multi-Dimensional Memory Frame possessed by both the self and the other individual. Conversely, in the case of Self–VR interaction, the generation and recognition paths on the system side are different from those on the human side. In the generation path, symbols defined in the system are converted into objects that can be perceived by the user. In the recognition path, human-generated objects are input to the system via sensors and converted into symbols that the system can handle. Both conversions are performed by programs implemented in the system.

In a VR environment, the system takes in the information emitted by human users and then determines the response to it. In any case, the input is represented as a symbol $\tilde{\mathbf{C}}$. Within the system, after setting a symbol $\bar{\tilde{\mathbf{C}}}$ to be transmitted in the next cycle of interaction, the symbol–object transformation is performed and the object $\bar{\tilde{\mathbf{O}}}_{\text{V}}(\bar{\tilde{\mathbf{C}}})$ is output to be perceived by the user. This generation path is denoted as [G-VV].

The user perceives the object $\bar{\tilde{\mathbf{O}}}_{\text{V}}(\bar{\tilde{\mathbf{C}}})$ and recognizes it as a symbol along the recognition path [R-VS]. Let $\hat{\mathbf{C}}'$ be the recognized symbol. The user activates his/her Multi-Dimensional Memory Frame along the generation path [G-SS] for $\hat{\mathbf{C}}'$ and obtains the corresponding object $\hat{\mathbf{O}}_{\mathbf{S}}(\hat{\mathbf{C}}')$. If the relation $\hat{\mathbf{O}}_{\mathbf{S}}(\hat{\mathbf{C}}') \equiv (\text{or} \approx)\ \bar{\tilde{\mathbf{O}}}_{\text{V}}(\bar{\tilde{\mathbf{C}}})$ holds, the interaction will proceed smoothly. If not, it will fail.

## IV. Method of $\bar{\mathbf{O}}_{\mathbf{V}}$ Generation with Symbol Grounding Secured

Only if the object generation path [G-VV] in the system is executed according to the user's meme mapping $[\cdots]_{\text{M-S}}$, that is, if the meme mapping in the system is based on $[\cdots]_{\text{M-AI}}$, it is possible to proceed with an interaction that guarantees symbol grounding. Since memes are knowledge that are passed down from generation to generation, they can be represented by symbols. This section proposes a method to implement meme mapping in the system. To this end, we first explain the process by which new symbols are assigned to patterns activated within the Multi-Dimensional Memory

2025, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

Frame during interaction with the external world. These patterns are then repeatedly used in communication with others—a process through which they become memes. Finally, based on this process, we add explanations for the A-, B-, and C-memes introduced in Section II-B2, and suggest a method for externalizing them by observing users' daily behavior.

### A. Memeification of Patterns

Figure 8 depicts how memes propagate in the reality field [15]. The process of propagation is facilitated by "symbolization." Using Figure 8, we will look at the process of meme propagation, focusing on the process of symbolization.

*1) Resonance with Objects in the Environment:* Suppose that Object-$O_1$, Object-$O_2$, or Object-$O_3$ appear in the environment (bottom-left in Figure 8). These objects are elements of $\mathbf{O} \ni \{\hat{\mathbf{O}}_\mathbf{S}, \bar{\mathbf{O}}_\mathbf{R}, \bar{\mathbf{O}}_\mathbf{V}\}$ shown in Figure 1. Here, it does not matter how the objects were generated. It is simply a situation in which they appear in front of self. Each resonates with the patterns of elements in the Perceptual-Multi-Dimensional Memory Frame of the self to activate them. This process is depicted as P-Resonance in Figure 2. Subsequently, activation propagates through the Multi-Dimensional Memory Frame through chain firings in the memory to activate the Relation-, Behavior-, and Motor-Multi-Dimensional Memory Frame. The activated portion of memory, which is shown as Pattern-$P_a$ at the bottom of Figure 8, will be used to select next actions in System 1 and System 2 cognitive processes through C-Resonance as shown in Figure 2. Pattern-$P_a$ is associated with different types of patterns, namely, patterns of elements of objects in the Perceptual-Multi-Dimensional Memory Frame, patterns of relations in the Relation-Multi-Dimensional Memory Frame, or patterns of temporal changes in the Behavior- and Motor-Multi-Dimensional Memory Frame.

*2) Mapping Environment on Brain:* The entities in the environment that resonate with Pattern-$P_a$, i.e., Object-$O \ni \{\text{Object-}O_1, \text{Object-}O_2, \text{Object-}O_3, \cdots\}$, are taken as a single object, Object-$O[P_a]$, in the action selection process conducted by System 1 and System 2 as shown at the bottom-right of Figure 8.

*3) Mapping Patterns on Word-Multi-Dimensional Memory Frame (Symbolization):* Once a unique name, Symbol-$S_a$, is attached to Pattern-$P_a$, Symbol-$S_a$ is stored in the Word-Multi-Dimensional Memory Frame while maintaining the relationship with Pattern-$P_a$. In Figure 8 this is shown by a thick double-headed arrow connecting bottom-right and top-left areas as mapping of pattern on the Word-Multi-Dimensional Memory Frame. Symbol-$S_a$ might be associated with any of the elements in Object-$O$, i.e., $\{\text{Object-}O_1, \text{Object-}O_2, \text{Object-}O_3, \cdots\}$. It is the process by which a large amount of information in a pattern is aggregated into a single symbol. By aggregating more information, the range of what can be expressed by a symbol is expanded [16].

*4) Importing* Symbol-$S_a$ *from the Word-Multi-Dimensional Memory Frame to Environment:* Once Pattern-$P_a$ is symbolized as Symbol-$S_a$ in the self's Word-Multi-Dimensional Memory Frame, it becomes a part of the environment, which is shown as an arrow from top-left to bottom-left of Figure 8. Namely, the relationship, Symbol-$S_a \in \mathbf{O}$, is established. This makes possible the following situation to emerge; when Symbol-$S_a$ appears as an object in the environment, it resonates with the Perceptual-Multi-Dimensional Memory Frame via P-Resonance to activate Pattern-$P_a$, followed by activation of patterns of relation in the Relation-Multi-Dimensional Memory Frame and temporal changes in the Behavior- and Motor-Multi-Dimensional Memory Frame to cause C-Resonance. In Section III-A, on the assumption that the the concept, $\hat{\mathbf{C}}$, exists, the generation path of the object representing it, $\hat{\mathbf{O}}_\mathbf{S}(\hat{\mathbf{C}})$, is illustrated (see Figure 3 and equation (1)). Alternatively, in this section, when an object $\mathbf{O}(*)$ of unknown generator with no symbol attached within self appears in the environment, a new symbol Symbol-$S_a$ is assigned to $\mathbf{O}(*)$. As a consequence, when the self recognizes the symbol in the environment which is available to anyone who is in it, the object, $\hat{\mathbf{O}}_\mathbf{S}(\text{Symbol-}S_a) \equiv \mathbf{O}(*)$, can be generated along the generation path shown by equation (1). Since Symbol-$S_a$ is created by starting from the generator unknown object, $\mathbf{O}(*)$, object $\mathbf{O}(*)$ could be recognized as Symbol-$S_a$. The symbol Symbol-$S_a$ is grounded on the real object $\hat{\mathbf{O}}_\mathbf{S}(\text{Symbol-}S_a) \equiv \mathbf{O}(*)$.

*5) Utilization of the Word-Multi-Dimensional Memory Frame in Language Behavior:* The processing flow explained so far is centered on Pattern-$P_a$ and therefore is closed in self and the environment. This situation corresponds to the one described in Section III-A. However, Symbol-$S_a$ incorporated in self's Word-Multi-Dimensional Memory Frame is available for interaction with others described in Section III-C, which is communication or language behavior depicted in top-right of Figure 8. Language behavior emerges via successive activation of the Word-, Relation-, Behavior-, and Motor-Multi-Dimensional Memory Frame, and utilization of them by System 1 and System 2 via C-Resonance. The resonated part of memory corresponds to Object-$O[P_a]$, which is the mapping of Pattern-$P_a$, generated via P-Resonance with the environment, to the brain.

In the situation where Symbol-$S_a$ incorporated in the self's Word-Multi-Dimensional Memory Frame, which is denoted as Symbol-$S_a$(self) to indicate explicitly it belongs to the self, is utilized in communication with others, Symbol-$S_a$(self) must connect to Symbol-$S_a$(other) incorporated in the other's Word-Multi-Dimensional Memory Frame. These symbols are apparently identical and refer to a unique symbol AbstractSymbol-$AS_a$. In Figure 8, the relationship between AbstractSymbol-$AS_a$ used in communication and Object-$O[P_a]$ directly connected with the self's symbol Symbol-$S_a$(self) is indicated by a thick double-headed arrow from bottom-right to top-right.

*6) Maintenance and Inheritance of Abstract Concepts:* The single abstract concept, AbstractSymbol-$AS_a$, is referred to Symbol-$S_a$(self) in the self's Word-Multi-Dimensional Memory Frame and Symbol-$S_a$(other) in the other's Word-Multi-Dimensional Memory Frame, which are denoted as $\hat{\mathbf{C}}$ and $\bar{\mathbf{C}}$, respectively, in Section III-C. Only when the meme mappings, $[\cdots]_{\text{M-S}}$ and $[\cdots]_{\text{M-O}}$, intervening in the respective symbol

2025, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

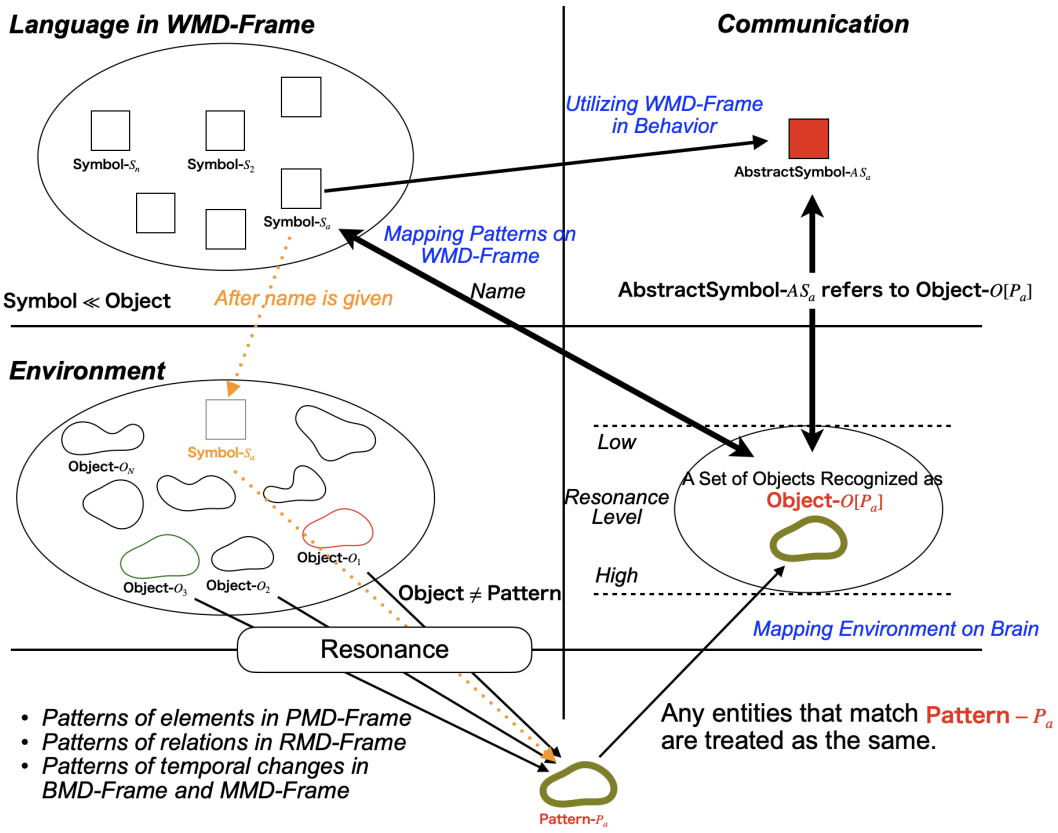


Figure 8. Memeification of patterns ([15, Figure 4] modified)

groundings of self and other are congruent, communication through AbstractSymbol-$AS_a \equiv \{\hat{\mathbf{C}}, \bar{\mathbf{C}}\}$ can be established while maintaining a common symbol grounding for both parties. When communication proceeds in this way, abstract concept AbstractSymbol-$AS_*$ is maintained and inherited. This is the reality of the meme. It is reasonable to assume that the higher the resonance level of Symbol-$S_a$ becomes, the longer AbstractSymbol-$AS_a$ is maintained and inherited as an effective communication medium.

### B. *Getting Memes into the Brain*

As explained in Section IV-A, a new symbol is generated within the Word-Multi-Dimensional Memory Frame. It becomes grounded as a real-world element through its association with a set of real-world objects and circulates within the behavioral ecology—including self and others—through interpersonal communication. It is a necessary condition for circulation that the symbols used in the community have the same symbol grounding. Those that circulate within the behavioral ecology are called memes. As each individual grows, the content of memes that are acquired changes. As explained in Section II-B2, first, the action-level memes involved in the basic action are acquired. Next, memes at the behavior level, which are sequences of actions, are acquired. Finally, the cultural-level meme—forming the background that shapes and supports actions and behaviors—is acquired.

Figure 9 shows an enlarged version of Pattern-$P_a$ displayed at the bottom center of Figure 8. The processes shown in Figure 8 as "Mapping Environment on Brain" and "Mapping Patterns on Word-Multi-Dimensional Memory Frame" are shown in the lower right and upper right of Figure 9, respectively. The process of "Mapping Patterns on Word-Multi-Dimensional Memory Frame" can be subdivided into the three types described above according to the developmental stage, as shown in Figure 9, depending on the degree of complexity of the mapping [13].

*1) Action-Level Memes:* During the period from birth to two to three years of age, humans generate several synapses that connect neural circuits in the brain and take in as much information as possible. The rate of synapse generation is then reduced, and the distribution of information up to that point is used to determine the basic characteristics of the sensory organs. Simultaneously, by initiating body movements and imitating the movements of the individuals around them, they acquire body movements that have been formed empirically and accumulatively as individual ecology. This is formed through life's skillful method of adjusting the growth of muscles and other parts of the body to external constraints. Simultaneously,

2025, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

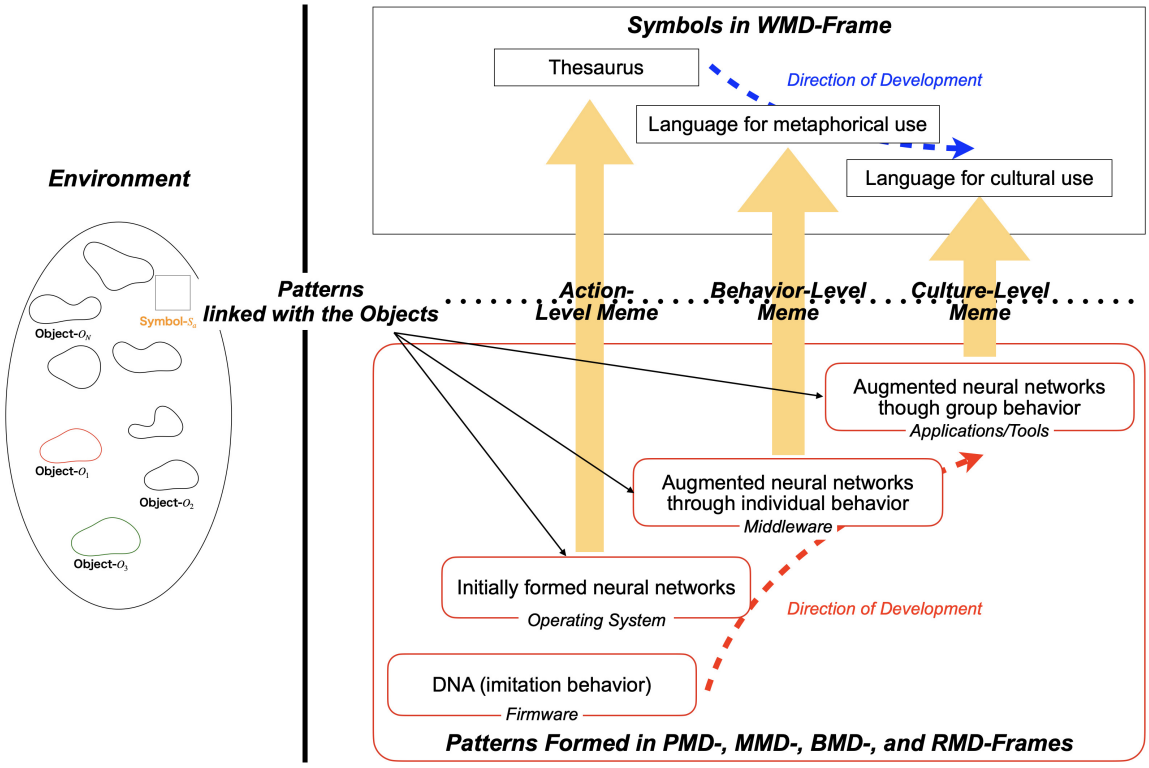


Figure 9. Structure of meme (adapted from [15, Figure 5])

information from the sensory organs is linked to bodily actions. The most important bodily functions formed at this stage are the voice and hand functions.

*2) Behavior-Level Memes:* Later, the voice paves the way to speech; the hands pave the way to tool use. Through continuous imitation, humans learn to use the words and tools of those around them. At this time, humans acquire a new hierarchy of actions by organizing and summarizing the fact that a particular collection of sounds evokes a particular response, and that the feel of a hand experienced through tactile sensation and the movement of a tool perceived visually are captured as an unified whole via the tool. This is achieved by linking the A-memes formed on the brain circuit according to the situation in which they are used, and making them available as a coherent whole.

*3) Culture-Level Memes:* Furthermore, words pave the way to language, and tools pave the way to the use of more complex machines. At this stage, humans learn to act as members of the culture and civilization of the group to which they belong, through imitation and through the experience of autonomous activities as members of the group. At this time, the B-memes is extended to be used in a complex manner, and culture-specific behavior patterns are formed.

### *C. Utilizing Memes in Behavioral Ecology with Realtime Constraints*

Figure 3 shows the situation in which symbol grounding is realized. This includes the process of starting with the concept, initiating the activated action, and the generation of object. However, this process can occur without exchanging the Perceptual-Multi-Dimensional Memory Frame and Word-, Relation-, and Behavior-Multi-Dimensional Memory Frame shown in Figure 3—that is, without ensuring existence through a link to perceptual reality. In practice, generating objects while ensuring their existence is inefficient and may therefore be omitted in real-life situations [12]. Objects that are guaranteed to be symbol grounded are often used in communication without explicitly verifying their grounding—particularly in interactions constrained by real-time demands. Under this situation, the meme mapping shown in equation (2) can be rewritten as follows:

$$\hat{\mathbf{C}} \Rightarrow \left[ \frac{\text{C-memes} \rightleftharpoons \text{B-memes}}{\textit{no reference to P-MDMF}} \rightleftharpoons \text{A-memes} \right]_{\text{M-S}} \Rightarrow \hat{\mathbf{O}}_{\mathbf{S}}(\hat{\mathbf{C}}) \quad (5)$$

### *D. Extraction of Memes through Mapping Memes into Information Systems*

Since memes are symbols, they can be expressed by language. Any elements appearing in the equation (5) are observable in the language. It is possible to extract A-, B-, and C-memes by organizing the verbal descriptions of actions performed top-down in daily life at the three levels of memes without referring to the Perceptual-Multi-Dimensional Memory Frame.

The mechanism by which the memes inherit information is

2025, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

analogous to an information system. This is shown in bottom-right of Figure 9. A-memes serve as the operating system that defines general patterns of spatial-temporal behavioral functions. B-memes represent middleware that extends the general patterns to concrete patterns. C-memes act as application tools that extend the concrete patterns to the ones that work in a number of groups of individuals.

The internal processing of information systems is represented by symbols. A necessary condition for the system and the user to achieve common symbol grounding is that $[\cdots]_{\text{M-AI}} \equiv [\cdots]_{\text{M-S}}$ holds. When trying to describe the memes contained in the contents of $[\cdots]_{\text{M-S}}$ in equation (5), the fact that an analogy can be made between memes at each level and the system mechanism indicates that the approach of describing memes by mapping them to information systems is effective. Representation of memes in the behavioral ecology in which the VR system is implemented regarding information system's symbols could be transferred to representation of $[\cdots]_{\text{M-AI}}$ that is equivalent to $[\cdots]_{\text{M-S}}$, opening the way for system implementation with guaranteed symbol grounding.

### E. Extraction of Memes: Examples

We have attempted to understand people's behavioral selection in various situations by assuming that it is performed according to MHP/RT, in the form of "under such-and-such a circumstance, such-and-such a person with such-and-such an attribute will express such-and-such a behavior." How a person acts reflects the functioning of the PCM process, and what the person does—the specific content of their actions—relates to the memes stored in their Multi-Dimensional Memory Frame and those activated during the execution of the actions. Cognitive Chrono-Ethnography (CCE) has been developed as a methodology for understanding behavior using this approach and has been applied to various situations [17].

Meme extraction has been attempted in some studies focusing on particular behavioral domains;

1) Navigation behavior that allows elderly people to reach their destinations without getting lost while obtaining necessary information from information displays in unfamiliar train stations [6];
2) The proficient creative activity that is the emanation of memes as the traditional skill of the ceramic artist, including memes that form the traditional skill that is handed down [18];
3) Proficient piano players practicing in preparation for competition performance [19].

It is expected that memes involved in interaction within VR environments can also be elucidated using the two-wheeled approach defined by the theory—MHP/RT—and its practical methodology, CCE, by consulting previous studies. By representing the memes activated within the VR environment using the vocabulary of the information system, it becomes possible to simulate the symbol grounding performed by the user. This enables the construction of a VR environment that connects seamlessly with the user in a forward-looking manner.

## V. Conclusion

Whether in the real world or in VR, individual humans execute the actions they select by running their PCM processes in response to objects that appear in the interacting world. This study posits that smooth, seamless interaction requires shared object meanings among interacting agents. We explored the concept of "symbol grounding," as discussed in artificial intelligence, within the context of user interaction in VR. AI will not achieve symbol grounding unless the same structured memes as humans are put into AI. In this study, while demonstrating that memes are observable, a path to how this is possible was presented.

We argued that symbol grounding is realized when we perceive and recognize an object generated from a certain symbol and match it with that symbol. In the object generation process, A-, B-, and C-memes, are involved in the conversion from symbols to objects. Since memes are passed down from generation to generation, we suggested that they could be extracted by observation, referring to the previous studies [6], [18], [19]. These are based on the MHP/RT [6]–[8] and the SMT [13], [15]. The ability to evaluate objects that are presented when users interact with VR environments from the symbol grounding perspective is an important issue in the development of seamless VR environments.

## Acknowledgment

This work was supported by JSPS KAKENHI Grant Number 19K12246 / 19K12232 / 20H04290 / 22K12284 / 23K11334, and National University Management Reform Promotion Project. The authors would like to thank Editage (www.editage.com) for the English language editing.

## References

[1] M. Kitajima, M. Toyota, and K. T. Nakahira, "Addressing the Symbol Grounding Problem in VR", in *AIVR 2025 : The Second International Conference on Artificial Intelligence and Immersive Virtual Reality*, 2025, pp. 56–62.

[2] A. Newell, "Physical Symbol Systems", *Cognitive Science*, vol. 4, no. 2, pp. 135–183, 1980. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1207/s15516709cog0402_2.

[3] A. Newell, *Unified Theories of Cognition (The William James Lectures, 1987)*. Cambridge, MA: Harvard University Press, 1990.

[4] J. E. Laird, A. Newell, and P. S. Rosenbloom, "Soar: An architecture for general intelligence", *Artificial Intelligence*, vol. 33, pp. 1–64, 1987.

[5] S. Harnad, "The symbol grounding problem", *Physica D: Nonlinear Phenomena*, vol. 42, no. 1, pp. 335–346, 1990, [retrieved: 2, 2025], ISSN: 0167-2789. DOI: https://doi.org/10.1016/0167-2789(90)90087-6.

[6] M. Kitajima and M. Toyota, "Simulating navigation behaviour based on the architecture model Model Human Processor with Real-Time Constraints (MHP/RT)", *Behaviour & Information Technology*, vol. 31, no. 1, pp. 41–58, 2012. DOI: 10.1080/0144929X.2011.602427.

[7] M. Kitajima and M. Toyota, "Decision-making and action selection in Two Minds: An analysis based on Model Human Processor with Realtime Constraints (MHP/RT)", *Biologically Inspired Cognitive Architectures*, vol. 5, pp. 82–93, 2013.

[8] M. Kitajima, *Memory and Action Selection in Human-Machine Interaction*. Wiley-ISTE, 2016, ISBN: 9781848219274.

2025, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

[9] M. Kitajima *et al.*, "Basic Senses and Their Implications for Immersive Virtual Reality Design", in *AIVR 2024 : The First International Conference on Artificial Intelligence and Immersive Virtual Reality*, 2024, pp. 31–38.

[10] D. Kahneman, "A Perspective on Judgment and Choice", *American Psychologist*, vol. 58, no. 9, pp. 697–720, 2003.

[11] D. Kahneman, *Thinking, Fast and Slow*. New York, NY: Farrar, Straus and Giroux, 2011.

[12] M. Kitajima, M. Toyota, J. Dinet, and K. T. Nakahira, "Transforming Conscious Goals into Unconscious Actions in Real-world Interactions: Real-world Use of Behavioral Ecological Memes via GOMS", *International Journal On Advances in Intelligent Systems*, vol. 18, no. 3 & 4, in press, 2025.

[13] M. Toyota, M. Kitajima, and H. Shimada, "Structured Meme Theory: How is informational inheritance maintained?", in *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, 2008, p. 2288.

[14] D. C. Dennett, *From Bacteria to Bach and Back: The Evolution of Minds*. W W Norton & Co Inc, Feb. 2018.

[15] M. Kitajima, M. Toyota, and J. Dinet, "How Resonance Works for Development and Propagation of Memes", *International Journal On Advances in Systems and Measurements*, vol. 14, pp. 148–161, 2021.

[16] M. Kitajima *et al.*, "Language and Image in Behavioral Ecology", in *COGNITIVE 2022 : The Fourteenth International Conference on Advanced Cognitive Technologies and Applications*, 2022, pp. 1–10.

[17] M. Kitajima, "Cognitive Chrono-Ethnography (CCE): A Behavioral Study Methodology Underpinned by the Cognitive Architecture, MHP/RT", in *Proceedings of the 41st Annual Conference of the Cognitive Science Society*, Cognitive Science Society, 2019, pp. 55–56.

[18] M. Kitajima, M. Toyota, and J. Dinet, "Art and Brain with Kazuo Takiguchi", in *COGNITIVE 2023 : The Fifteenth International Conference on Advanced Cognitive Technologies and Applications*, 2023, pp. 1–10.

[19] K. T. Nakahira, M. Kitajima, and M. Toyota, "Understanding Practice Stages for a Proficient Piano Player to Complete a Piece: Focusing on the Interplay Between Conscious and Unconscious Processes", *International Journal on Advances in Life Sciences*, vol. 16, pp. 164–177, 2024.

2025, © Copyright by authors, Published under agreement with IARIA - www.iaria.org